

أخلاقيات الذكاء الاصطناعي للتنفيذيين

أبريل 2022



بسم الله الرحمن الرحيم



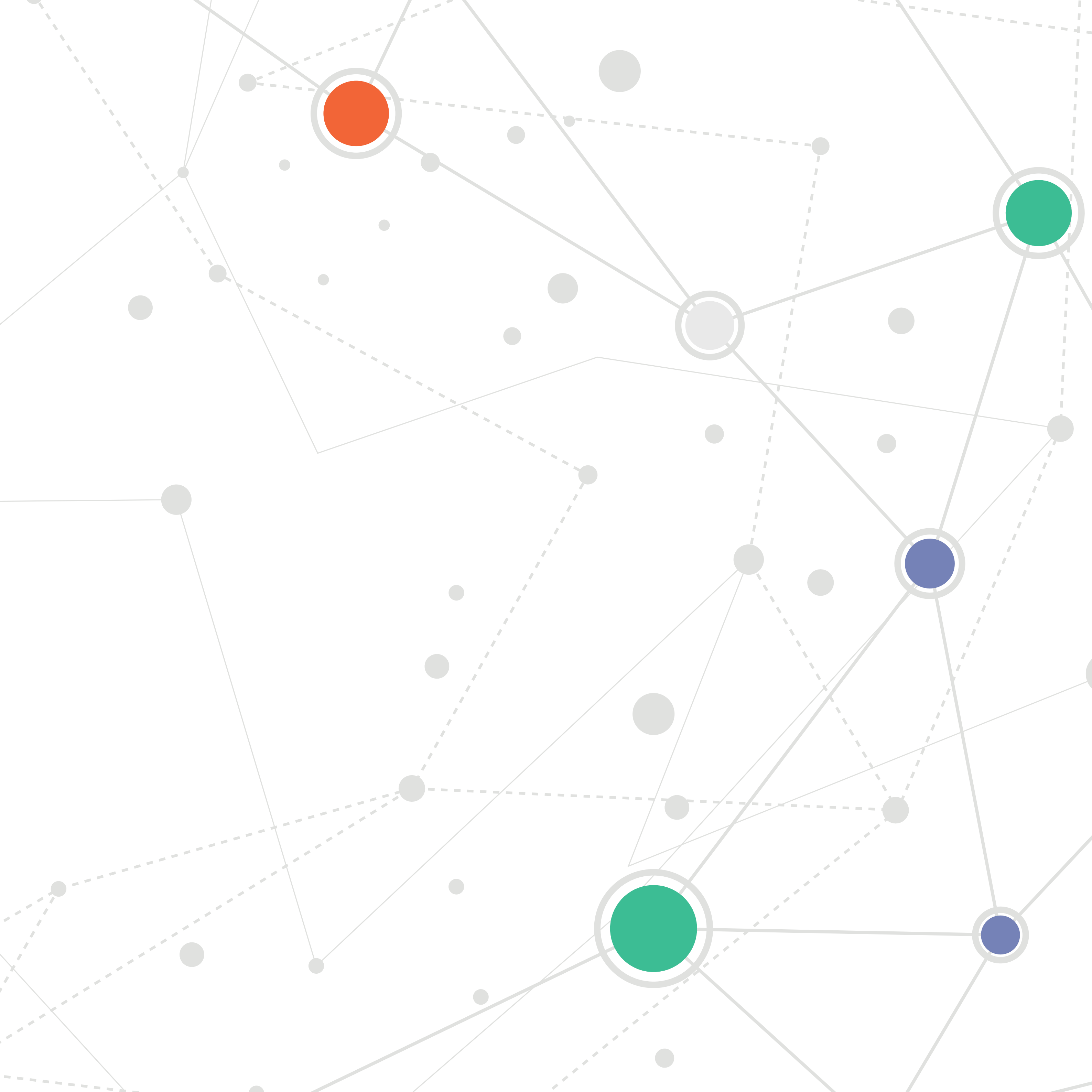


تمهيد

تتسابق كثير من الدول المتقدمة لتسخير الذكاء الاصطناعي في بناء اقتصادات متينة تعتمد على البيانات والتقنيات الحديثة بطريقة أخلاقية ومستدامة. وتعد المملكة العربية السعودية إحدى الدول السباقة في تسخير إمكاناتها نحو اقتصاد رقمي متنوع وقائم على المعرفة مع الالتزام بتعزيز المساعي الدولية لضمان الاستخدام الأمثل لتقنيات الذكاء الاصطناعي بما يعود بالنفع على البشرية كافة تحقيقاً لمستهدفات رؤية 2030م. ومن هذا المنطلق، أسهمت الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) في تعزيز مفهوم أخلاقيات الذكاء الاصطناعي على المستوى العالمي والمحلي عن طريق عدد من المبادرات الرائدة. فعلى المستوى العالمي، شاركت (سدايا) في تمثيل المملكة العربية السعودية لدى منظمة اليونسكو لمناقشة توصيات أخلاقيات الذكاء الاصطناعي. كما نظمت (سدايا) القمة العالمية للذكاء الاصطناعي تحت شعار "الذكاء الاصطناعي لخير البشرية" بهدف فتح آفاق الحوار حول مستقبل الذكاء الاصطناعي، وتعزيز الاستخدام الآمن والأخلاقي لما فيه مصلحة البشرية. وعلى المستوى المحلي، أطلقت (سدايا) مجموعة من السياسات التنظيمية لحوكمة البيانات الوطنية وضمان استخدامها بطريقة أخلاقية وآمنة تسهم في تعزيز الاقتصاد الوطني مع المحافظة على السرية والخصوصية. وهذا بفضل الله أولاً ثم بفضل الرؤية الحكيمة والقيادة الرشيدة والدعم المستمر من لدُن مولاي خادم الحرمين الشريفين وولي عهده الأمين — حفظهما الله — وهذا يحتم علينا مضافرة الجهود والسعي بجد لتمكين الذكاء الاصطناعي المسؤول ولتكون مملكتنا الحبيبة مثالاً يحتذى به إقليمياً وعالمياً.

الدكتور عبدالله بن شرف الغامدي

رئيس الهيئة السعودية للبيانات والذكاء الاصطناعي





محتويات





مقدمة

يشهد العالم اليوم تطوراً متسارعاً في الذكاء الاصطناعي وتقنياته، الأمر الذي أسهم في ظهور عدد من الحلول المبتكرة في مختلف المجالات التي من شأنها رفْع الأداء الاقتصادي وكفاءة الأعمال في كل من القطاعين الحكومي والخاص، وتحسين جودة حياة الأفراد والمجتمعات. وتشير التوقعات الى أن الذكاء الاصطناعي سيعزز الاقتصاد العالمي بنحو (14%) بحلول عام 2030م، أي ما يقارب (15.7) تريليون دولار أمريكي (حوالي 58.9 تريليون ريال سعودي)[1]، ومن المتوقع أن تتبنى (70%) من الشركات تقنيات الذكاء الاصطناعي بحلول عام 2030م[2].

ومع توسع انتشار هذه التقنيات زاد النقاش حول أخلاقيات الذكاء الاصطناعي واستخدامه المسؤول في مختلف المجالات، وظهرت عدة تحديات ومخاوف أثارت الشكوك والقلق حول تطوير الذكاء الاصطناعي وتبني تقنياته، مثل التحيز والتمييز وانتهاك حقوق الإنسان. ولذلك اتجهت عدة منظمات عالمية وقطاعات حكومية ومؤسسات بحثية وشركات تجارية إلى تحديد أهم المبادئ الأخلاقية والممارسات التي يمكن عن طريقها مواجهة هذه التحديات ومعالجتها، وتوقع المخاطر المستقبلية المحتملة، وضمان تطوير أنظمة ذكاء اصطناعي أخلاقية عادلة وآمنة.

يهدف هذا الدليل إلى تبسيط مفهوم أخلاقيات الذكاء الاصطناعي وتوضيح أهميتها، وتلخيص أهم مبادئها. كما يستعرض الدليل أهم الاعتبارات العامة وأفضل الممارسات التطبيقية التي ينصح بالأخذ بها عند تطوير أنظمة الذكاء الاصطناعي أو تبنيها، بالإضافة إلى مجموعة من الأدوات التي تساعد في تحقيق مبادئ أخلاقيات الذكاء الاصطناعي. كما يُشير الدليل إلى أبرز الجهود التي قدمتها (سدايا) في مجال أخلاقيات الذكاء الاصطناعي على المستوى المحلي والعالمي.

[1]Rao, D. & Verweij, G. Sizing the prize: What's the real value of AI for your business and how can you capitalise?. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf (2017).

[2]Bughin, J., Seong, J., Manyika, J., Chui, M. & Joshi, R. Notes from the AI frontier: Modeling the impact of AI on the world economy. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy (2018).

1. تعريف أخلاقيات الذكاء الاصطناعي

يمكن تعريف أخلاقيات الذكاء الاصطناعي بأنها:

"مجموعة من القيم والمبادئ والأساليب لتوجيه السلوك الأخلاقي في تطوير تقنيات الذكاء الاصطناعي واستخدامها."

وتهدف أخلاقيات الذكاء الاصطناعي إلى تحديد كل ما هو صائب أو خاطئ بما يتوافق مع سياسة المنظمات والقطاعات الحكومية والمؤسسات والشركات، وتقديم الإرشادات طوال دورة حياة أنظمة الذكاء الاصطناعي ابتداءً من البحث والتخطيط وحتى التطوير والاختبار والتشغيل.

أبعاد أخلاقيات الذكاء الاصطناعي الرئيسية

لأخلاقيات الذكاء الاصطناعي أربعة أبعاد رئيسية يجب مراعاتها منذ المراحل الأولى لتبني الحلول القائمة على الذكاء الاصطناعي أو تطويرها أو استخدامها، بهدف معرفة أين تنشأ القضايا والتحديات المختلفة المتعلقة بالأخلاقيات وكيفية التعامل معها أو تجنبها مستقبلاً:

المجتمع والسياسات التنظيمية

دراسة وضع الذكاء الاصطناعي في المجتمع وأثره المحتمل على الجوانب المختلفة كالتعليم والتوظيف والدخل وغيرها، مع الأخذ بعين الاعتبار مستوى قبوله، ومدى تطور السياسات والإستراتيجيات التنظيمية.

التنظيمات الداخلية

حوكمة السياسات والإجراءات الداخلية والمعايير المتعلقة بالذكاء الاصطناعي مع مراعاة أن ترتبط برؤية المنظمة الداخلية ومهامها وأهدافها ومتطلباتها القانونية والتنظيمية.

القوى العاملة

التقييم المستمر لمدى تأثير أنظمة الذكاء الاصطناعي على القوى العاملة والمسؤوليات التنظيمية، وتطوير المهارات لتتناسب مع المسؤوليات الجديدة، ورفع وعي جميع مستويات القوى العاملة وتشجيعهم على تبني الممارسات الأخلاقية عند تطوير أنظمة الذكاء الاصطناعي واستخدامها.

التقنية والبيانات

مراعاة الممارسات الأخلاقية خلال دورة حياة أنظمة الذكاء الاصطناعي، ويشمل ذلك طرق جمع البيانات وتصنيفها ومعالجتها واختيارها وحمايتها وكيفية بناء النماذج واختبارها، وكيفية استخدامها في البيئة التشغيلية.





2. أهمية أخلاقيات الذكاء الاصطناعي

تُشير نتائج عدد من الدراسات الاستطلاعية إلى إدراك متخذي القرار والتنفيذيين لأهمية أخلاقيات الذكاء الاصطناعي، وأنها ستكون جزءاً أساسياً في إستراتيجيات الأعمال المستقبلية. فقد أكد (78%) من كبار متخذي القرار في القطاع الخاص على أهمية أن تكون مخرجات أنظمة الذكاء الاصطناعي عادلة وآمنة وموثوقة[3]. ويعتقد (63%) من المديرين التنفيذيين أن أخلاقيات الذكاء الاصطناعي ستصبح عنصراً أساسياً في إستراتيجية شركاتهم خلال العامين القادمين[4]. ومع ذلك أبدى (50%) من المديرين التنفيذيين قلقهم بشأن المخاطر الأخلاقية للذكاء الاصطناعي ومدى تأثيرها في شركاتهم[5]، إذ يثير الذكاء الاصطناعي عدداً من التحديات والقضايا الأخلاقية التي يمكن أن تؤدي إلى انخفاض مستوى الثقة في تبني أنظمة الذكاء الاصطناعي، فضلاً عن الإضرار بالشركات وسمعتها والدخول في قضايا قانونية، مما قد يؤدي إلى تناقص القيمة السوقية للشركات أو تعرضها لخسائر مادية أو معنوية.

[3] AI Ethics From Roadblock to Scale: The Global Sprint Towards AI. https://filecache.mediaroom.com/mr5mr_ibmnews/183710/Roadblock-to-Scale-exec-summary.pdf (2020).

[4] State of Responsible AI: 2021. https://www.fico.com/en/latest-thinking/market-research/state-responsible-ai-2021 (2021).

[5] Thriving in the era of pervasive AI. https://www2.deloitte.com/cn/en/pages/about-deloitte/articles/state-of-ai-in-the-enterprise3-rd-edition.html (2020).

التحديات الأخلاقية



التأثير على الفرد والمجتمع

باتت أنظمة الذكاء الاصطناعي تؤثر بطريقة ما أو بأخرى على معظم جوانب الحياة، وقد يؤدي سوء تصميم هذه الأنظمة بطريقة مقصودة أو غير مقصودة، فضلاً عن إمكانية استخدامها بصورة سيئة، إلى التأثير سلبياً في الفرد والمجتمع على حد سواء.



تحديد المسؤولية والمساءلة

تثير أنظمة الذكاء الاصطناعي تساؤلات حول المسؤول عن مخرجاتها وما قد تسببه من أضرار وخسائر مادية ومعنوية. وتتمثل المشكلة في عدم وجود قوانين تنظم استخدامها أو تشريعات تحدد المسؤوليات وتحمي المستخدم عندما تكون النتائج غير متوقعة أو خاطئة.



عدم الشفافية وإمكانية التفسير

تتسم نماذج تعلم الآلة بقدرتها على توليد الارتباطات بين البيانات المدخلة والنتائج دون برمجة صريحة مما يجعلها معقدة وذات طبيعة أشبه بالصناديق السوداء التي لا يعرف ما بداخلها ولا يمكن تفسير قراراتها.



انتهاك الخصوصية

تَعتمد معظم أنظمة الذكاء الاصطناعي على البيانات التي قد تتضمن بيانات شخصية، ومع انتشار مصادر البيانات والتوسع في جمعها واستخدامها أو مشاركتها زاد خطر انتهاك حقوق الفرد وتهديد خصوصيته، خاصةً عندما يتم ذلك دون علمه أو موافقته.



التحيز

تفضيل نظام الذكاء الاصطناعي لمجموعات على أخرى بسبب عرق أو جنس أو غيره، فمثلاً يؤدي استخدام بيانات متحيزة في بناء نماذج الذكاء الاصطناعي وتدريبها إلى تحيّز مخرجاتها وقراراتها.

3. مبادئ أخلاقيات الذكاء الاصطناعي

تبذل بعض الحكومات والمؤسسات الجهود لوضع أطر قانونية وتنظيمية لتطوير الذكاء الاصطناعي واستخدامه بطريقة مسؤولة وأخلاقية، ونُشرت الكثير من الإرشادات والمبادئ الأخلاقية التي تجاوز عددها (173) وثيقة إرشادية[6]. ومع أن غالب هذه الإرشادات وما تتضمنه من مبادئ غير مُلزمة وتختلف من دولة إلى أخرى ومن قطاع إلى آخر، إلا أنها تُسهم بشكل كبير في وضع الأسس الرئيسية لتطبيق أخلاقيات الذكاء الاصطناعي، وتقدم توصيات مفيدة لصناع السياسات والتنفيذيين، وتعالج كثيراً من القضايا والتحديات الأخلاقية، وتغطي جوانب عدة من مراحل دورة حياة بناء أنظمة الذكاء الاصطناعي.

يركز هذا الدليل على خمسة مبادئ رئيسية لأخلاقيات الذكاء الاصطناعي، حسب ما تبنته منظمة التعاون والتنمية الاقتصادية (OECD)[7] عام 2019م، واعتمدتها دول مجموعة العشرين بالإضافة إلى (42) دولة، وأكدت عليها مؤخراً اتفاقية منظمة الأمم المتحدة للتربية والعلم والثقافة (اليونسكو) عام 2021م، وتبنتها (193) دولة[8]، من ضمنها المملكة العربية السعودية.



[6] AI Ethics Guidelines Global Inventory. https://inventory.algorithmwatch.org/ (2020).

[7] The OECD Artificial Intelligence (AI) Principles - OECD.AI. https://oecd.ai/en/ai-principles (2019).

[8] Report of the Social and Human Sciences Commission (SHS). https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14 (2021).



النمو الشامل والتنمية المستدامة والرفاهية

يهدف هذا المبدأ إلى تسخير تقنيات الذكاء الاصطناعي واستخدامها بطريقة مبتكرة تعود بالنفع على الأفراد والمجتمعات والبيئة وتسهم في النمو والازدهار والصالح العام وتعزز أهداف التنمية العالمية. ويشمل هذا المبدأ ما يلي:

- **النمو الشامل والرفاهية:** تقديم المنفعة وتعزيز السلام ورفاهية الإنسان وازدهاره، وخلق الفرص الاجتماعية والاقتصادية.
- **الاستدامة:** تقديم منفعة مستدامة تسهم في تحسين النظام البيئي وحمايته وحفظ مصادره.

الأهمية

تتمثل أهمية هذا المبدأ في الحاجة إلى تطوير أنظمة ذكاء اصطناعي واستخدامها لدعم مستقبل مستدام وتعزيز الأثر الإيجابي في الأفراد والمجتمعات على المدى القريب والبعيد. ويمكن تلخيص تأثير أنظمة الذكاء الاصطناعي في النقاط الآتية:

المجتمع

يؤدي الاعتماد على أنظمة الذكاء الاصطناعي في مجالات الحياة المختلفة إلى تغيير المفاهيم الاجتماعية وانفصال الأفراد عن محيطهم الاجتماعي، مما ينتج عنه جمود العلاقات الإنسانية وفقدان المهارات الاجتماعية التي بدورها تؤثر سلباً في صحة الإنسان.



البيئة

يتطلب تدريب أنظمة الذكاء الاصطناعي على البيانات الضخمة استهلاكاً كبيراً للطاقة ينتج عنه تلوث كربوني هائل، إذ أشارت دراسة إلى أن مقدار انبعاثات الكربون من تدريب نموذج واحد في معالجة اللغات الطبيعية قد يعادل الانبعاثات الناتجة عن صنع خمس سيارات واستخدامها[9].



الاقتصاد

يؤثر تبني تقنيات الذكاء الاصطناعي في حجم الوظائف ونوعيتها وفرص العمل المتاحة، إذ سيلغي الذكاء الاصطناعي بعضها وسيخلق فرصاً أخرى تتطلب مهارات وكفاءات مختلفة، مما سيؤدي إلى انخفاض الأمان الوظيفي والتوزيع غير المتكافئ للدخل.

[9] Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. https://aclanthology.org/P1355-19.pdf

اعتبارات عامة

تحقيق الفائدة للأفراد والمجتمعات والبيئة طوال دورة حياة نظام الذكاء الاصطناعي.

مراعاة الاستدامة والمسؤولية البيئية وإفادة البشرية بما في ذلك الأجيال القادمة عند تطوير نظام الذكاء الاصطناعي أو استخدامه.

وضوح آثار تطوير نظام الذكاء الاصطناعي واستخداماته الإيجابية والسلبية المباشرة وغير المباشرة على الأفراد والمجتمع والبيئة.

ممارسات تطبيقية

المشاركة المسبقة

إشراك أصحاب المصلحة بصورة مسبقة في إدارة الاستخدام المسؤول للذكاء الاصطناعي سعياً نحو تحقيق نتائج مفيدة ومستدامة للإنسان والبيئة عن طريق تنمية القدرات البشرية وتعزيز الإبداع ودراسة الآثار الاقتصادية والاجتماعية وغيرها.

رفع الوعي

تثقيف مصممي نظام الذكاء الاصطناعي ومطوريه حول التأثيرات الإيجابية والسلبية التي يمكن أن يحدثها هذا النظام في كل من الأفراد والمؤسسات والمجتمعات.

المراقبة والتقييم

تحديد آليات لمراقبة آثار نظام الذكاء الاصطناعي وتقييمها طوال دورة حياته.

الإفصاح عن الآثار المحتملة

الكشف عن التأثيرات المتوقعة من الحلول القائمة على الذكاء الاصطناعي في المستخدم داخل المنظمة وخارجها.

إشراك الموظفين في عملية اتخاذ القرار

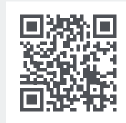
إعطاء الموظفين المتأثرين بأنظمة الذكاء الاصطناعي داخل المنظمة الفرصة للمشاركة في عملية اتخاذ القرار عند تبني هذه الأنظمة أو تطويرها.

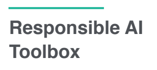


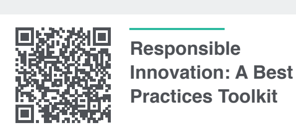
تشجيع البحث العلمي

تعزيز جوانب البحث والتطوير داخل المنظمة وذلك لتطوير أنظمة ذكاء اصطناعي تساعد في معالجة المجالات ذات الاهتمام العالمي كأهداف التنمية المستدامة.

أدوات مساعدة









القيم الإنسانية والعدالة

يركز هذا المبدأ على تسخير تقنيات الذكاء الاصطناعي وتصميمها واستخدامها بطريقة تحترم القانون وحقوق الإنسان والقيم الإنسانية بما فيها المساواة والتنوع، بما يضمن للأفراد والمجتمعات العدالة والإنصاف. وتتمحور قيم هذا المبدأ حول محورين أساسيين هما:

- **الإنسان:** أن تكون أنظمة الذكاء الاصطناعي تخدم البشرية وتراعي قضايا حقوق الإنسان.
- **العدالة:** أن يكون نظام الذكاء الاصطناعي طوال دورة حياته عادلاً بما يضمن التوزيع العادل والمتكافئ للفوائد والفرص، بالإضافة إلى خلو النظام من التحيز والتمييز. وتركز العدالة على أربعة أبعاد، هي:

البيانات
أن تكون البيانات المستخدمة في تدريب أنظمة الذكاء الاصطناعي محايدة وممثلة بصورة صحيحة ودقيقة وشاملة وقابلة للتعميم وتحقق الأهداف المرجوة منها.

التصميم
خلو تصاميم نماذج الذكاء الاصطناعي من المتغيرات أو العمليات أو الاستدلالات أو الارتباطات غير المبررة أو المرفوضة من الناحية الأخلاقية.



البناء
أن تكون الأنظمة مطورة بطريقة مسؤولة وخالية من التحيز.

النتيجة
خلو النتيجة من الآثار الغير عادلة على حياة الأفراد.



الأهمية

تتمثل أهمية مبدأ القيم الإنسانية والعدالة في مواءمة أنظمة الذكاء الاصطناعي طوال دورة حياتها مع القيم الإنسانية كالحريات والمساواة والعدالة الاجتماعية وحماية البيانات والخصوصية، فضلاً عن حقوق المستخدمين والعدالة التجارية، إذ تعتمد أنظمة الذكاء الاصطناعي في تصميمها وتطويرها بشكل رئيسي على العنصر البشري الذي يفرض أفكاره ومفاهيمه على خوارزميات الذكاء الاصطناعي، ولذا من الممكن أن تغذى هذه الأنظمة في مراحل دورة حياتها بالأخطاء البشرية والتحيزات ابتداءً من استخراج البيانات وجمعها ومعالجتها وحتى مراحل بناء النماذج وتطبيقها. ومن أبرز أنواع التحيز ما يلي:



تحيز ضمني

يحدث هذا النوع من التحيز بوعي أو بغير وعي عند الاعتماد على التجارب والمفاهيم الشخصية وإغفال العوامل الأخرى المؤثرة عند تطوير نماذج الذكاء الاصطناعي.

تحيز النموذج

يقصد به انحياز خوارزمية الذكاء الاصطناعي عند تصنيف البيانات، مما يؤدي إلى ضعف القدرة التنبؤية أو الفشل في التعميم على البيانات الجديدة.



تحيز البيانات

يقصد به تحيز بيانات التدريب نحو مجموعة أو أكثر دون الأخرى، نتيجة لكون البيانات غير شاملة ولا تمثل التوزيع الحقيقي.

اعتبارات عامة



ممارسات تطبيقية



مراعاة القوانين

وضع آليات تطوير متوافقة مع القوانين والقيم الإنسانية والعدالة والالتزام بها عند تطوير نظام الذكاء الاصطناعي واستخدامه.



الموازنة بين الفوائد والمخاطر

توضيح الفوائد والمخاطر المتوقعة لنظام الذكاء الاصطناعي على حقوق الإنسان مع الحفاظ على الموازنة بينهما والأخذ بعين الاعتبار الاتجاهات والتغيرات المستقبلية.



بناء فرق متنوعة

مراعاة تكوين فرق تطوير من مجموعة متنوعة من الأعمار والخلفيات والثقافات عند تطوير نظام الذكاء الاصطناعي.



تحديد آليات لتقييم مدى التمييز

استخدام المقاييس كنسبة التأثير السلبي، أو التأثير الهامشي، أو فرق المتوسط لفهم متى وكيف يتسبب نظام الذكاء الاصطناعي في ممارسات تمييزية.



تطوير منهجيات لجمع البيانات وتصنيفها

الإشراف على عمليات إعداد وتطوير أساليب جمع البيانات وتصنيفها ومعالجتها، ووضع القيود والمتطلبات بطريقة واضحة لضمان خلو البيانات من التحيز والتمييز.



التحقق من جودة البيانات

استخدام منهج معياري واضح لفحص البيانات المستخدمة في تدريب نظام الذكاء الاصطناعي في وقت مبكر وبصورة متكررة للتأكد من صحة البيانات وشموليتها وتنوعها وخلوها من التحيز.



مراعاة معايير سهولة الوصول

التصميم الشامل لنظام الذكاء الاصطناعي وإتاحة إمكانية الوصول للنظام والبيانات للجميع بغض النظر عن عمرهم أو جنسهم أو قدراتهم أو خصائصهم كالأشخاص ذوي الاحتياجات الخاصة.



تقييم موثوقية المخرجات

اختبار مخرجات نظام الذكاء الاصطناعي بالنسبة لأصحاب المصلحة أو المجموعات التي قد تتأثر سلباً للتأكد من دقة نتائج النظام وموثوقيتها.



إتاحة حق إدارة البيانات

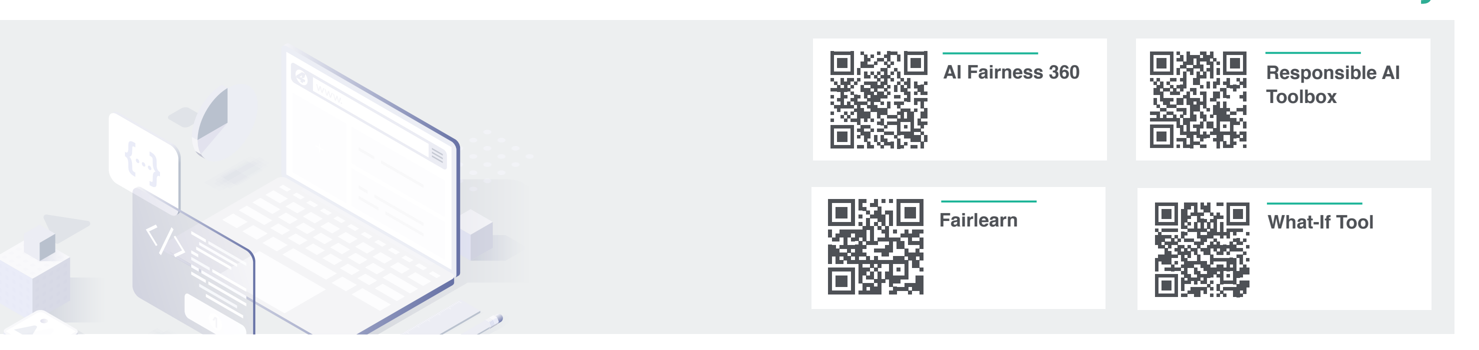
إعطاء المستخدمين الحق في إدارة بياناتهم المستخدمة في تدريب نظام الذكاء الاصطناعي.



وضع آلية للتغذية الراجعة

تطوير آلية للحوار المفتوح مع المستخدمين بهدف الكشف عن التحيزات أو التحديات التي يواجهها المستخدم.

أدوات مساعدة



الشفافية وقابلية التفسير

يُعد مبدأ الشفافية وقابلية التفسير الركيزة الأخلاقية الأساسية المرتبطة بفهم وشرح أنظمة الذكاء الاصطناعي ومخرجاتها، إذ تسمح لأصحاب المصلحة بفهم المراحل الرئيسية للتطوير وعمليات اتخاذ القرار. ويتكون هذا المبدأ من محورين رئيسيين:

- **قابلية التفسير:** شرح كيفية وصول أنظمة الذكاء الاصطناعي إلى قراراتها ومخرجاتها بطريقة بسيطة ومفهومة.
- **الشفافية:** فهم كيفية تنفيذ كل مرحلة من مراحل دورة حياة نظام الذكاء الاصطناعي لتقديم معلومات تتعلق بما يلي:
  - حقيقة استخدام نظام الذكاء الاصطناعي في عمليات صنع القرار.
  - أي معلومات تتعلق بالاحتمالات أو المنطق المعتمد في النظام.
  - مجموعات البيانات التي يستخدمها النظام.
  - الغرض من النظام وكيفية استخدامه.

الأهمية

تعود أهمية مبدأ الشفافية وقابلية التفسير إلى كونها طريقة لتقليل الضرر وتحسين الذكاء الاصطناعي وتعزيز ثقة المستخدمين، إذ يُنظر إلى معظم أنظمة الذكاء الاصطناعي وخاصة نماذج تعلم الآلة على أنها صناديق سوداء لا يمكن فهمها أو شرح ما يحدث بداخلها أو تفسير كيفية وصول هذه الأنظمة إلى قراراتها ومخرجاتها، مما يؤدي إلى مجموعة من التحديات، من ضمنها:



التحيز

قد يؤدي التعتيم في أنظمة الذكاء الاصطناعي وعدم توضيح نوع البيانات المستخدمة في التدريب أو كيفية اتخاذها للقرارات والعوامل التي أدت إلى توصياتها النهائية إلى اتخاذ قرارات خاطئة ومتحيزة نحو عرق أو جنس أو عمر ما.

قلة ثقة المستخدمين

تعد ثقة المستخدمين عاملاً مهماً ويؤدي انخفاضها إلى قلة تبني أنظمة الذكاء الاصطناعي، إذ إن المستخدم لا يمكن أن يثق في أنظمة الذكاء الاصطناعي ما دام أنه لم يفهم كيفية عملها.



المخاطر القانونية والأمنية

قد ينتج عن الغموض في أنظمة الذكاء الاصطناعي صعوبة فهم العمليات الداخلية مما يؤدي إلى تحدياتٍ في تحديد المسؤولية أو توزيعها عند حدوث خطأ ما في النظام.

اعتبارات عامة



مراعاة تعزيز الشفافية والقابلية للتفسير عند تطوير نظام الذكاء الاصطناعي ووضع معايير التقييم.

قابلية توضيح قرارات نظام الذكاء الاصطناعي ومنهجياته بطريقة يسهل فهمها.



إمكانية تتبع العوامل الرئيسية التي أثرت على قرارات نظام الذكاء الاصطناعي.

إدراك المستخدم مشاركة نظام ذكاء اصطناعي في عملية صنع القرار أو عند التفاعل معه.

ممارسات تطبيقية



توثيق المعلومات والعمليات

الاحتفاظ بالمعلومات خلال دورة حياة نظام الذكاء الاصطناعي لفترة زمنية معينة تتناسب مع نوع القرار أو مجال النظام، على أن تشمل المعلومات ما يلي:

- مصادر بيانات التدريب وطرق جمعها ومعالجتها، وكيفية نقلها وحفظها، مع توضيح التدابير المتخذة للحفاظ على دقتها وخصوصيتها مع مرور الوقت.
- تصميم نموذج الذكاء الاصطناعي والخوارزميات المستخدمة.
- التغيرات التي أجريت على النظام مع تحديد المسؤول عن كل تغيير.
- سجل مخرجات عملية اتخاذ القرار للتحقق منها عند الحاجة.



وضع المعايير

تطوير معايير لقياس مستوى الشفافية متناسبة مع غرض نظام الذكاء الاصطناعي ومراجعتها بصورة مستمرة واستخدامها لتقييم النظام بموضوعية أثناء التطوير واقتراح آليات لتحسين مستوى الشفافية.



استخدام الشفافية والقابلية للشرح كمعيار

تشجيع فريق مطوري أنظمة الذكاء الاصطناعي على استخدام الشفافية والقابلية للشرح كمعيار عند اختيار الخوارزميات أو النماذج.



تفسير القرارات

توضيح مصدر القرارات وكيف أسهم نظام الذكاء الاصطناعي في إصدارها جزئياً أو كلياً.



استخدام لغة غير تقنية

توضيح كل ما يتعلق بنظام الذكاء الاصطناعي بلغة مفهومة لأصحاب المصلحة وتشمل:

- البيانات التي يستخدمها النظام.
- الفئات التي تتأثر بالنظام.
- أهم العوامل التي تؤثر في القرارات الناتجة عنه.



توفير وسيلة لتتبع القرار

تقديم وسيلة سهلة الوصول لتفسير «رحلة اتخاذ القرار» ويمكن عن طريقها تتبع قرارات نظام الذكاء الاصطناعي بطريقة سهلة الاستخدام وواضحة.



إيضاح المقصد العام

شرح المقصد العام للنظام أو أكثر العوامل المرجح إسهامها في اتخاذ القرار عندما تكون أنظمة الذكاء الاصطناعي غير قابلة للتفسير لأسباب تعود إلى الملكية الفكرية أو غيرها.



مراعاة الخصوصية

الإشعار المسبق للأفراد في حال استخدام بياناتهم لتطوير نظام ذكاء اصطناعي.



الاتصال الفاعل

التعبير بوضوح عن فوائد نظام الذكاء الاصطناعي ومخاطره المحتملة، بالإضافة إلى الإجراءات المتخذة لتجنب هذه المخاطر مع مراعاة الاختلافات الاجتماعية والثقافية للفئات المستهدفة.



الإفصاح بالهوية الآلية

التوضيح للمستخدم حقيقة تفاعله مع نظام الذكاء الاصطناعي مع تحديد مدى هذا التفاعل كما في بوت المحادثة (Chatbot).



توفير آلية للاستفسار

تمكين المستخدم من التساؤل حول مخرجات نظام الذكاء الاصطناعي.



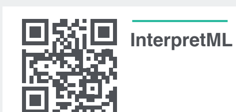
وضع آلية للتغذية الراجعة

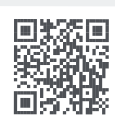
جمع المعلومات والآراء من أصحاب المصلحة الخارجيين والداخليين حول نظام الذكاء الاصطناعي ومدى موثوقيته.



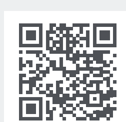
أدوات مساعدة

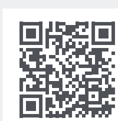

Responsible AI Toolbox


InterpretML


AI FactSheets 360


AI Explainability 360


Model Card Toolkit


Explainable AI



المتانة والأمن والسلامة

يهتم هذا المبدأ بمعالجة تحديات أمن أنظمة الذكاء الاصطناعي وسلامتها ومواجهة المخاطر الأمنية والتغلب عليها طوال دورة حياتها، إضافة إلى سلامة مخرجات أنظمة الذكاء الاصطناعي في جميع الظروف المختلفة. ويتضمن هذا المبدأ ركيزتين رئيسيتين هما:

- **المتانة:** إنتاج مخرجات ذات موثوقية عالية في ظل الظروف المختلفة.
- **الأمن والسلامة:** وتشمل جانبين:

الجانب الرقمي

يُعنى بسلامة البيانات وجودتها وحمايتها واحترام مبادئ الخصوصية، وكذلك حماية أنظمة الذكاء الاصطناعي وصحة نتائجها وقراراتها.

الجانب المادي

خلو أنظمة الذكاء الاصطناعي من المخاطر التي تؤثر في السلامة المادية وضمان عدم إلحاقها الضرر بالأفراد بطريقة مقصودة أو غير مقصودة.

الأهمية

تتمثل أهمية مبدأ المتانة والأمن والسلامة في حماية أنظمة الذكاء الاصطناعي ومخرجاتها، وكذلك سلامة البيانات في ظل الهجمات العدائية التي تهدف في الغالب إلى إضعاف أداء النماذج أو التلاعب في نتائجها، ومن هذه الهجمات ما يلي:



استخراج النموذج

محاولة سرقة النموذج أو البيانات الحساسة والوصول إلى التفاصيل الداخلية لنظام الذكاء الاصطناعي عبر إعادة بناء النموذج أو استخراج البيانات المستخدمة في التدريب.



هجمات التهرب

التلاعب بالمدخلات أثناء استخدام نظام الذكاء الاصطناعي دون التأثير في بيانات التدريب لتفادي اكتشاف أنها ضارة، كاستخدام مقاييس حيوية مزيفة في أنظمة التحقق من الهوية.

تسمم البيانات

التلاعب ببيانات التدريب أو تغييرها لتكون غير مفيدة أو مؤذية أو مسماة بصورة غير صحيحة.



اعتبارات عامة



ضمان عمل نظام الذكاء الاصطناعي بطريقة موثوقة وآمنة في جميع الظروف.

ضمان أمن وسلامة جميع الأفراد المشغلين أو المستخدمين أو غيرهم.

الاعتماد على تدابير أمن وسلامة تتناسب مع حجم المخاطر المحتملة.



ضمان احترام خصوصية البيانات وحمايتها.

ممارسات تطبيقية



توثيق المعلومات والعمليات

الاحتفاظ بالمعلومات خلال دورة حياة نظام الذكاء الاصطناعي لفترة زمنية معينة تتناسب مع نوع القرار أو مجال النظام، على أن تشمل المعلومات ما يلي:

- توافق النظام واستخدامه مع الهدف المقصود منه.
- مجموعات البيانات المستخدمة والعمليات المتبعة.
- دقة نتائج النظام وقدرته على إجراء تنبؤات وقرارات صحيحة.



تقييم درجة التعميم

اختبار سلوك نظام الذكاء الاصطناعي وقدرته على تقديم تنبؤات صحيحة بناءً على بيانات جديدة في عدد من السيناريوهات لقياس مدى قدرته على العمل بصورة سليمة عند اختلاف المدخلات أو الظروف أو عند تكرار التجارب ذاتها.

وضع منهجية إدارة المخاطر

تحديد إطار عمل لإدارة المخاطر المحتملة وتطبيقه طوال دورة حياة نظام الذكاء الاصطناعي، على أن تشمل المنهجية الخطوات الآتية:

- **تحديد المخاطر المحتملة**

  دراسة جميع المخاطر المتوقع حدوثها سواء أكانت نتيجة لتطوير النظام أم استخدامه بأي صورة كانت مقصودة أو غير مقصودة، بالإضافة إلى الهجمات التي قد تخل بأمن البيانات وسلامتها أو قد تؤثر سلباً في سلوك النظام ونتائجه.

- **تقييم المخاطر المحتملة**

  تحليل المخاطر والعوامل وتحديد أولوياتها حسب خطورتها ومدى احتمالية وقوعها.

- **تحديد طرق الاستجابة**

  وضع إجراءات للتعامل مع هذه المخاطر والاستجابة لها حال حدوثها سواءً عبر حل المشكلة ذاتها أو التخفيف من آثارها.

- **توفير آليات للإبلاغ والتحقيق**

  تحديد القنوات والإجراءات للإبلاغ أو التحقيق عند حدوث أي من المخاطر المحددة وتوفير المساعدة في الوقت المناسب.



تحديد آليات الحماية

وضع الآليات المناسبة لحماية نظام الذكاء الاصطناعي من الثغرات التي يمكن استغلالها لتسميم البيانات أو التأثير في النماذج أو البنية التحتية واتخاذ الخطوات المناسبة لمنع الهجمات العدائية.



توثيق العمليات مع الصيانة المستمرة

التأكد من جودة البيانات وسلامتها عبر توثيق جميع العمليات ومجموعات البيانات المستخدمة واختبارها في كل خطوة من مراحل دورة حياة نظام الذكاء الاصطناعي مع صيانة البيانات وتصحيحها بصورة مستمرة.



تحديد صلاحيات الوصول للبيانات

وضع البروتوكولات المناسبة لتحديد الأفراد المؤهلين للوصول إلى البيانات داخل المنظمة والظروف التي تمكّنهم من ذلك.



الامتثال للتشريعات

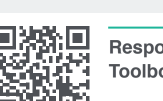
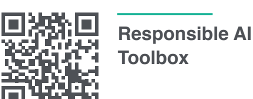
مراعاة التشريعات الوطنية المتعلقة بالبيانات سعياً إلى لوصول إلى الاستخدام المسؤول والأخلاقي للبيانات.

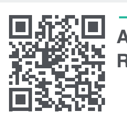


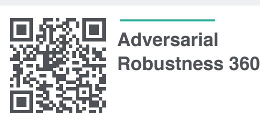
وضع المقاييس

تحديد المقاييس الملائمة لمبدأ المتانة والأمن والسلامة لتحديد انتهاكات الخصوصية المحتملة مثلاً في نظام الذكاء الاصطناعي والحرص على اختبار هذه المعايير طوال دورة حياة النظام.









المسؤولية والمساءلة

يركز هذا المبدأ على ضمان نزاهة أنظمة الذكاء الاصطناعي وعدالتها وتوضيح المسؤوليات والصلاحيات والأطر التنظيمية المتبعة لجميع الإجراءات أو القرارات خلال دورة تطوير هذه الأنظمة أو نشرها واستخدامها، ويتضمن هذا المبدأ ما يلي:

- **المسؤولية:** ما يلزم المسؤول ذو الأهلية من واجبات وما يترتب عليها من تبعات في جميع مراحل دورة تطوير أنظمة الذكاء الاصطناعي واستخدامه.
- **المساءلة:** قدرة المسؤول على تفسير قرارات نظام الذكاء الاصطناعي ونتائجه الواقعة ضمن مسؤوليته.

الأهمية

تعود أهمية مبدأ المسؤولية والمساءلة إلى حاجة الأفراد والمؤسسات لتوضيح حدود المسؤولية والمساءلة عند تطوير أنظمة الذكاء الاصطناعي واستخدامها بهدف معالجة القضايا القانونية وتجاوز الفجوة التي تحدثها أنظمة الذكاء الاصطناعي في النظام القانوني بسبب استبدال الأنظمة الذكية بالعامل البشري في المجالات الآتية:



تحديد المتسبب

صعوبة تحديد المسؤول عند حدوث مشكلة ما.

تحقيق العدالة

صعوبة تحديد العقوبات والأحكام القانونية.



التعويض

صعوبة مطالبة الضحايا بالتعويضات عن طريق النظام القانوني.

اعتبارات عامة



قابلية تحديد المسؤولين — سواء كانوا افرداً أو كيانات تنظيمية — عن المراحل المختلفة لدورة حياة نظام الذكاء الاصطناعي.

توزيع المساءلة عن نتائج نظام الذكاء الاصطناعي على المسؤولين بناء على أدوارهم في جميع مراحل دورة حياة النظام.

ألاّ تُنسب المسؤولية عن الأداء أو الخسائر أو الأضرار الناتجة في جميع مراحل دورة حياة نظام الذكاء الاصطناعي إلى النظام بحد ذاته.



إتاحة الفرصة للمستخدمين للاعتراض على قرارات نظام الذكاء الاصطناعي ومخرجاته.



ألاّ تُصدر أنظمة الذكاء الاصطناعي قرارات مهمة بالنيابة عن المستخدمين دون الحصول على موافقتهم المسبقة.



ألا يُترك أمر اتخاذ القرارات المصيرية كالحياة أو الموت لنُظم الذكاء الاصطناعي.

ممارسات تطبيقية



الاستعانة بالبحوث العلمية

تطوير أنظمة ذكاء اصطناعي قائمة على البحوث العلمية والتطبيقية والأدلة والبراهين، والتشجيع على تبني هذه الأنظمة.

توثيق عمليات التطوير

حفظ عمليات تطوير أنظمة الذكاء الاصطناعي في سجلات مفصلة مع تحديد المسؤول قانونياً عن كل منها لتشمل ما يلي:

- الغرض المقصود من النظام
- بيانات التدريب
- بيئة التدريب
- مصادر البيانات
- الخوارزميات
- الرسوم البيانية
- مميزات النموذج
- واجهات المستخدم
- النتائج



مراعاة القوانين واللوائح

فهم التشريعات والسياسات ذات الصلة والامتثال لها مع الحرص على تضمينها عند تطوير أنظمة الذكاء الاصطناعي واستخدامها.



رفع الوعي

الاهتمام بتوفير المعلومات للمهندسين والمطورين والمديرين المعنيين بمنتجات الذكاء الاصطناعي وتدريبهم حول القضايا الأخلاقية والمعايير المختلفة الخاصة بأنظمة الذكاء الاصطناعي.



تقييم المخاطر المحتملة

دراسة الآثار المباشرة وغير المباشرة التي قد تنتج عن نظام الذكاء الاصطناعي على المستخدمين ومراجعة مدى توافقها مع المبادئ والأخلاقيات داخلياً وخارجياً.



توضيح سياسة المنظمة

مراعاة وضوح السياسات الخاصة بالمنظمة حول قضايا المسؤولية والمساءلة الداخلية لجميع فرق العمل من المصممين والمطورين وغيرهم.



توضيح حدود المسؤولية

تحديد أين تنتهي مسؤولية المنظمة أو القائمين على نظام الذكاء الاصطناعي وتوضيحها للمستهلك النهائي.



تحديد الخطط البديلة

وضع آليات للاستئناف وخطط الطوارئ وتمكين الإشراف البشري على مخرجات نظام الذكاء الاصطناعي لمعالجة مخرجات النظام الخاطئة أو التخفيف منها طوال دورة حياة النظام.



تخصيص لجنة لمعالجة الشكاوي

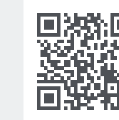
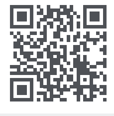
تعيين أفراد مسؤولين للنظر في الشكاوى والتحقيق في أسباب الضرر الناتج عن تطوير نظام الذكاء الاصطناعي أو استخدامه ومعالجة هذه الأسباب.



تحديد آليات التعويض

وضع إجراءات خاصة بالتعويض يمكن عن طريقها تعويض المستخدمين المتأثرين بقرارات نظام الذكاء الاصطناعي ونتائجه الخاطئة.

أدوات مساعدة


Responsible AI Toolbox

Algorithmic Accountability Policy Toolkit

4. جهود سدايا في أخلاقيات الذكاء الاصطناعي

في ظل سعي المملكة العربية السعودية نحو الريادة في الذكاء الاصطناعي مع المحافظة على السيادة الوطنية الرقمية على البيانات، أطلقت المملكة متمثلة في (سدايا) عدة أنظمة وسياسات تنظم جمع البيانات الشخصية ومعالجتها ومشاركتها، وتضمن المحافظة على خصوصية أصحاب هذه البيانات وحماية حقوقهم، كما تعمل على تعزيز أداء المؤسسات ورفع مستوى شفافيتها ومسؤوليتها.

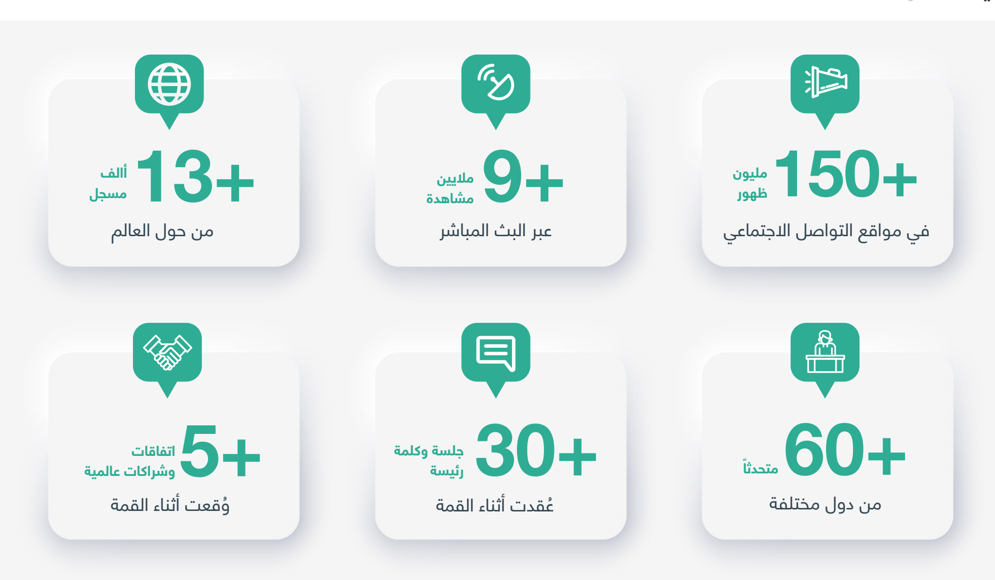
على المستوى العالمي

المشاركة في إعداد توصيات اليونسكو بشأن أخلاقيات الذكاء الاصطناعي

شاركت (سدايا) في تمثيل المملكة العربية السعودية لدى اليونسكو لمناقشة توصيات أخلاقيات الذكاء الاصطناعي والتي تم اعتمادها مؤخراً في نوفمبر 2021م، وتهدف هذه التوصيات إلى إيجاد إطار عالمي للقيم والمبادئ والإجراءات اللازمة لإرشاد الدول فيما يخص وضع تشريعاتها أو سياساتها الأخرى المتعلقة بالذكاء الاصطناعي بما يتوافق مع القانون الدولي.

القمة العالمية للذكاء الاصطناعي

نظمت (سدايا) النسخة الأولى من القمة العالمية للذكاء الاصطناعي في عام 2020م تحت شعار "الذكاء الاصطناعي لخير البشرية" بهدف مشاركة الأفكار والآراء من صناع القرار والمبتكرين والخبراء والمستثمرين من مختلف أنحاء العالم وفتح أبواب التعاون العالمي لصالح النهوض بالمجال.





على المستوى المحلي

مكاتب إدارة البيانات

أسست (سدايا) **(137) مكتباً مختصاً** في حوكمة البيانات وإدارتها في الجهات الحكومية لتطبيق الأنظمة والمعايير والسياسات المتعلقة بإدارة البيانات وحماية البيانات الشخصية ومتابعة مدى الالتزام بها.

الأنظمة

أصدرت (سدايا) في عام 2021م **نظام حماية البيانات الشخصية** لضمان حق أصحاب البيانات الشخصية في الاطلاع على بياناتهم ومعرفة الغرض من جمعها ومعالجتها، بالإضافة إلى ضمان حقهم في الوصول إلى البيانات وطلب تصحيحها أو تحديثها أو إتلافها عند انتهاء الغرض من جمعها.

السياسات التنظيمية

تعمل (سدايا) على بناء قواعد تشريعية متينة لتمكين تقنيات البيانات والذكاء الاصطناعي، وعليه طورت **ثمان سياسات** في مجال البيانات.

سياسة تصنيف البيانات

تهدف إلى وضع إطار موحد لتصنيف البيانات بناءً على نتائج تقييم الأثر المترتب عن الإفصاح عنها أو عن محتواها، ويتضمن التصنيف أربعة مستويات: سري للغاية، وسري، ومقيد، وعام، مع تحديد الضوابط المناسبة لكل مستوى على حدة. والأخذ بعين الاعتبار مبادئ الوصول، والاستخدام، والاحتفاظ بالبيانات وأرشفتها، ومشاركتها والتخلص منها.

سياسة حماية البيانات الشخصية

تهدف إلى وضع الأحكام والإجراءات التي تنظم جمع البيانات الشخصية ومعالجتها للحفاظ على خصوصية أصحابها وحماية حقوقهم. وتحقق هذه السياسة مبادئ المسؤولية والشفافية وأمن البيانات وجودتها، والمراقبة والامتثال.

سياسة مشاركة البيانات

تهدف إلى تعزيز ثقافة مشاركة البيانات وضمان تحقيق التكامل بين الجهات الحكومية والحصول على البيانات من مصادرها الصحيحة، والحد من ازدواجيتها وتعارضها وتعدد مصادرها. وتحقق هذه السياسة مبادئ تعزيز ثقافة المشاركة، والشفافية، والمسؤولية المشتركة وأمن البيانات والاستخدام الأخلاقي لها.

سياسة حرية المعلومات

تهدف إلى وضع الأحكام والإجراءات التي تنظم ممارسة حق الوصول إلى المعلومات العامة أو الحصول عليها، بما يضمن تعزيز مبادئ الشفافية في جميع الجهات الحكومية، والمساواة، والإفصاح عن المعلومات العامة.

سياسة البيانات المفتوحة

تهدف إلى وضع قواعد عامة تنظم إتاحة البيانات غير المصنفة على إحدى درجات السرية ونشرها بصورة استباقية تشجيعاً للبحث والابتكار، وتطويراً لنموذج الحوكمة وإشراك الجميع، ودعماً للنمو الاقتصادي، وتحقيقاً لمبدأ الشمولية، وعدم التمييز.

سياسة حماية البيانات الشخصية للأطفال ومن في حكمهم

تهدف إلى حماية الأطفال على شبكة الإنترنت، ومساعدة الجهات المختصة على حمايتهم من مخاطر جمع ومعالجة بياناتهم الشخصية. وتحقق هذه السياسية مبادئ تعزيز حماية البيانات الشخصية للأطفال، وتحسين المحتوى الرقمي.

القواعد العامة لنقل البيانات الشخصية خارج الحدود الجغرافية للمملكة

تهدف إلى تنظيم نقل البيانات الشخصية خارج الحدود الجغرافية للمملكة لضمان المحافظة على السيادة الوطنية على هذه البيانات والخصوصية والحماية لأصحابها وتعزيز مبادئ حماية البيانات الشخصية والخصوصية.

ضوابط إدارة البيانات الوطنية وحوكمتها وحماية البيانات الشخصية

تهدف إلى تعزيز البيانات وتنميتها كأصول وطنية عن طريق إدارتها وتمكينها ورفع القيمة المضافة منها وتحقيق مبادئ تنمية البيانات وتكاملها وتعزيز جودتها.



مراجع

1. A Framework for the Ethical use of Advanced Data Science Methods in the Humanitarian Sector. https://5f2cd2ba741-c4-b-29ae00-47a8291b1d3c.filesusr.com/ugd/d1cf5c_6af8feb771194453817d62c92cee2a21.pdf (2020).
2. AI ethics: A business imperative for boards and C-suites. https://www2.deloitte.com/us/en/pages/regulatory/articles/ai-ethics-responsible-ai-governance.html (2021).
3. AI Ethics. https://www.adobe.com/sa_en/about-adobe/aiethics.html (2021).
4. Ammanath, B. & Blackman, R. Everyone in Your Organization Needs to Understand AI Ethics. https://hbr.org/07/2021/everyone-in-your-organization-needs-to-understand-ai-ethics (2021).
5. A practical guide to Responsible Artificial Intelligence (AI). https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf (2019).
6. Australia's Artificial Intelligence Ethics Framework. https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework (2021).
7. Avin, S. et al. Filling gaps in trustworthy development of AI. https://www.science.org/doi/10.1126/science.abi7176 (2021).
8. Burkhardt, R., Hohn, N. & Wigley, C. Leading your organization to responsible AI. https://www.mckinsey.com/~/media/mckinsey/business20%functions/mckinsey20%analytics/our20%insights/leading20%your20%organization20%to20%responsible20%ai/leading-your-organization-to-responsible-ai.pdf?shouldIndex=false (2019).
9. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf (2019).
10. Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2019).

11. Everyday Ethics for Artificial Intelligence. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (2019).

12. Explainable AI.https://www.ibm.com/sa-en/watson/explainable-ai (2021).

13. From Principles to Practice An interdisciplinary framework to operationalise AI ethics. https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf (2020).

14. Golbin, I. & Luciana Axente, M. 9 ethical AI principles for organizations to follow. https://www.weforum.org/agenda/06/2021/ethical-principles-for-ai/ (2021).

15. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. https://www.nature.com/articles/s42256-019-0088-2 (2019).

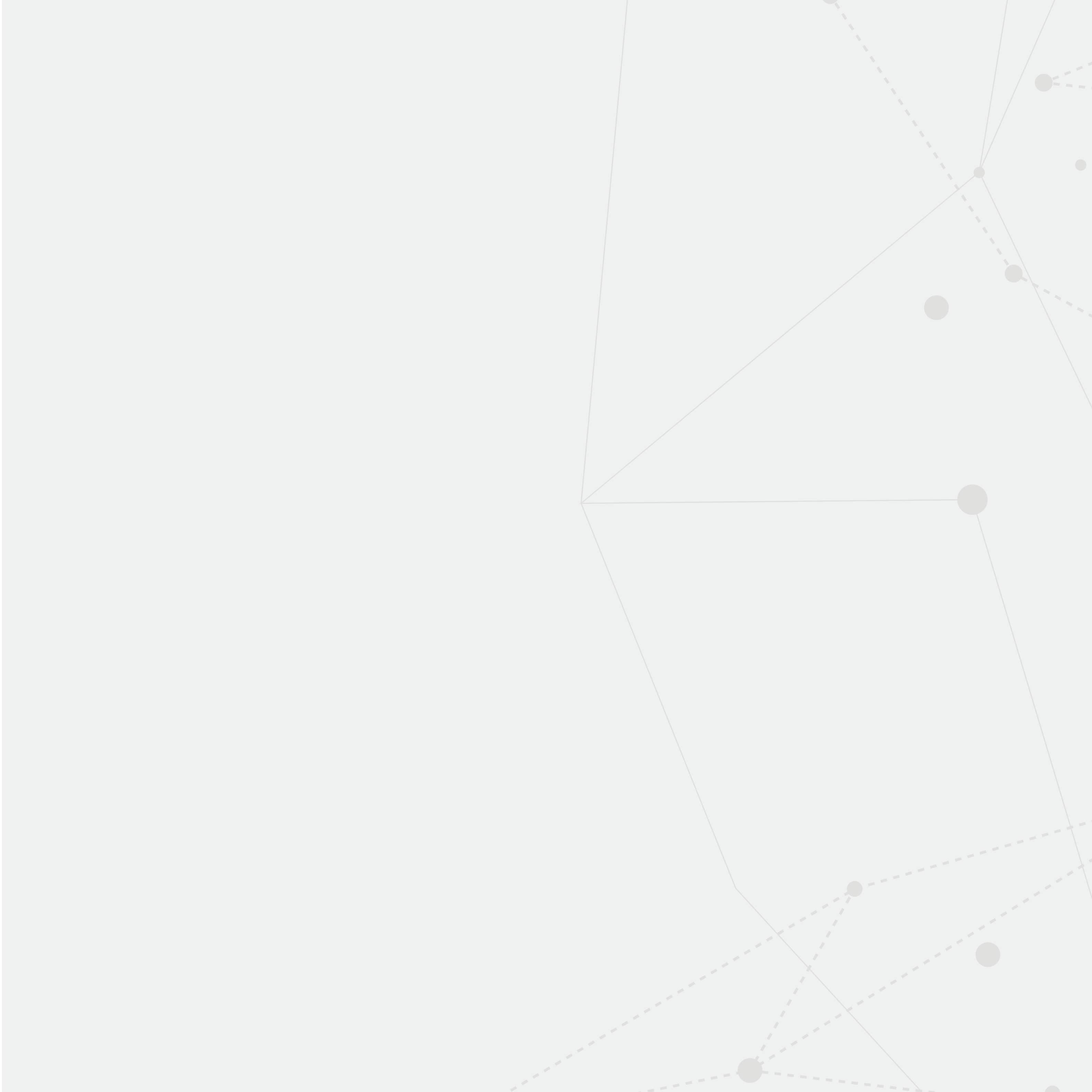
16. Leslie, D. Understanding artificial intelligence ethics and safety. https://www.turing.ac.uk/sites/default/files/06-2019/understanding_artificial_intelligence_ethics_and_safety.pdf (2019).

17. OECD AI Policy Observatory. https://www.oecd.ai/ (2021).

18. Responsible AI #AIForAll. http://www.niti.gov.in/sites/default/files/02-2021/Responsible-AI22022021-.pdf (2021).

19. Thieullent, A. et al. AI and the Ethical Conundrum: How organizations can build ethically robust AI systems and gain trust. https://www.capgemini.com/wp-content/uploads/10/2020/AI-and-the-Ethical-Conundrum-Report.pdf (2020).

20. Understanding artificial intelligence ethics and safety. https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety#understanding-what-ai-ethics-is (2019).



سلسلة الأدلة الإرشادية





